EXHIBIT 22

From: Simon, Jeremy <jeremy_simon@med.unc.edu>
Sent: Wednesday, October 26, 2022 5:29 PM
To: Pearson, Brandon L.
Subject: Re: FW: [EXTERNAL] Human Genetics and Genomics Advances Decision HGG-ADVANCES-D-22-00031

Okay next time it's ready for review, you can recommend Albert Keung at NC State as a reviewer. His lab works on genomics in neurodevelopment, organoids, single cell, UBE3A, etc and I think he understands the nuance but importance of work like this

Jeremy

(Sent remotely, please excuse any terseness)

On Oct 13, 2022 8:07 AM, "Pearson, Brandon L." <blp2125@cumc.columbia.edu> wrote:

From: Chung, Wendy K. <wkc15@cumc.columbia.edu>
Sent: Friday, June 10, 2022 6:48 AM
To: Pearson, Brandon L. <blp2125@cumc.columbia.edu>; Baker, Brennan H. <bhb2128@cumc.columbia.edu>
Subject: RE: [EXTERNAL] Human Genetics and Genomics Advances Decision HGG-ADVANCES-D-22-00031

Brandon,

Yes, reviewer 2 is a stickler.
If helpful, I'm happy to have you present your work at our Friday lab meeting if helpful to talk/think through your results/future experiments together and help improve the rigor of the analyses. You may already know how to do that. I'd like you to present at my group separately to help me think through analyzing a few cases in SPARK that have a particularly high de novo mutation frequency. We can also do that separately if you like.

I don't know the journals related to mutagenesis/environmental exposures to consider, but I suggest getting away from the hard core genetic journals. I think it will be hard to convince those reviewers/editors of the importance of this work without some really striking results.

Don't give up. New ideas are harder to change minds but long term are of high impact.

Wendy

From: Pearson, Brandon L. <blp2125@cumc.columbia.edu>
Sent: Thursday, June 9, 2022 10:32 PM
To: Baker, Brennan H. <bhb2128@cumc.columbia.edu>; Chung, Wendy K. <wkc15@cumc.columbia.edu>
Subject: Fwd: [EXTERNAL] Human Genetics and Genomics Advances Decision HGG-ADVANCES-D-22-00031

Hi

No luck with the mutation paper. I have to see the humor in this as the first reviewer says the results are to be perfectly expected and intuitive whereas the second reviewer says we went about it all wrong. Wendy knew that the message would be hard to impress on genetics reviewers. Let's use the technical and statistical issues that'll were raised here and improve it and find another place to send it.

Best

Brandon

Begin forwarded message:

> **From:** HGG Advances <em@editorialmanager.com>
> **Date:** June 9, 2022 at 19:52:23 MDT
> **To:** "Pearson, Brandon L." <blp2125@cumc.columbia.edu>
> **Subject: [EXTERNAL] Human Genetics and Genomics Advances Decision HGG-ADVANCES-D-22-00031**
> **Reply-To:** HGG Advances <hggadvances@ashg.org>
>
> CC: jxchong@uw.edu, mbamshad@uw.edu, mbamshad@u.washington.edu
>
> Jun 09, 2022
>
> Dear Dr. Pearson,
>
> Thank you for sending your manuscript titled "Environmental carcinogens disproportionally mutate genes implicated in neurodevelopmental disorders" (HGG-ADVANCES-D-22-00031) to Human Genetics and Genomics Advances. We have received the reports of the reviewers and we enclose the comments they have provided for transmission to the authors. After careful consideration, we have decided that your paper is not a good fit for Human Genetics and Genomics Advances.
>
> It is important to stress that, because of the limited number of articles that Human Genetics and Genomics Advances can publish, many of the papers sent out for evaluation are ultimately not accepted. The decision not to accept a paper that undergoes detailed review is based not only on the specific technical limitations detailed in the comments to the authors, but also on the overall priority awarded the submission based on novelty and scope, a recommendation that is communicated directly to us. For this reason, a revised paper addressing only the technical concerns of the referees is not likely to change the decision against acceptance, and we believe it is best for the paper to be submitted elsewhere. We nevertheless hope that you will not hesitate to submit future manuscripts to Human Genetics and Genomics Advances.

Sincerely,

Alexis Battle
Associate Editor, Human Genetics and Genomics Advances

Michael J. Bamshad, M.D.
Editor-in-Chief, Human Genetics and Genomics Advances

HGGAdvances@ashg.org
919-650-1459

Reviewer #1: In "Environmental carcinogens disproportionately mutate genes implicated in neurodevelopmental disorders", Baker et al. analyze the distributions of mutations observed in iPSCs exposed to various chemicals and conclude that NDD-associated genes are more likely than randomly selected genes to harbor exposure-associated mutations. They further examine potential predictors of mutational rate probabilities and conclude that gene length and GC content are the major drivers of the NDD-gene-mutation-enrichment.

At a biological level, the findings are not surprising, as the major conclusion is to recapitulate the well-established fact that NDD genes tend to be longer than non-NDD genes and, as an effectively trivial result, will naturally tend to accumulate more mutations by essentially any mutational process. In fact, I would argue that the title is misleading, in that NDD genes do not appear to be disproportionately sensitive to chemical mutagenesis once one accounts for the fact that they are longer; i.e., the main result is consistent with a simple null model that gene-length predicts gene-mutation-accumulation. Relatedly, the statement towards the end that "environmentally induced mutations may play a greater role in neurodevelopmental disease than previously assumed" is also not supported, in that any "previously assumed" role would necessarily be a function of gene length (at the very least).

At a technical level, while the analyses appear basically sound, they tend to ignore a lot of well-established knowledge regarding mutation rates. For example, as CpG to TpG transitions are ~10X more likely to occur than other mutations, dinucleotide composition should be explicitly accounted for in all the analyses presented (note that this is likely to make a simple "null model" of mutation-probability prediction even stronger); similarly, it may be useful to incorporate experimentally measured iPSC methylation data as an additional predictor. Further, richer mutational models, such as one in which transitions and transversions are considered separately, would likely also be informative; in fact one might imagine that Ti-to-Tv ratios will vary considerably among differing chemical exposures, and differ from naturally arising DNA polymerase errors, and potentially allow for observations about chemical mutagenesis that are not simply

a function of length. Finally,
additional analyses to consider would be comparisons with gene-specific de novo mutational rate predictions, such as those from Samocha et al, and gene-specific "tolerance" metrics built on empirically defined metrics of population variation accumulation, such as those from Petrovski et al. These types of analysis might be capable of revealing differences between patterns of chemical mutagenesis and those observed via basic mutation and selection models.

In the absence of a new biological finding, which would require substantially more detailed assessments that account for the rich knowledgebase of mutation rate drivers, I would not recommend proceeding with publication.

Reviewer #2: HGG-ADVANCES-D-22-00031
Environmental carcinogens disproportionally mutate genes implicated in neurodevelopmental disorders.
The authors evaluated if 769 disease genes were likely to be affected by mutagens. They used WGS from one iPSC clonal line exposed to 12 classes of environmental carcinogens and evaluated their mutational rate. With this, the authors identify a set of genes susceptible to mutations that they further investigate.

I find this manuscript poorly written with an absolute lack of justification and support for its methods, analyses and conclusions. First, the authors assume the reader have extensive and detailed knowledge of the original Kubac et al paper. Moreover, the authors barely explain their methods, making impossible for anyone to understand their work. Attempting to reproduce this work is beyond anyone. It is not possible to evaluate if their conclusions are supported by their analyses as it is.

Specific comments:
1) The authors used somatic mutations called using WGS mapped to the Human reference genome GRCh37/hg19 (as indicated by Kubac et al), however their gene length and location was done using GENCODE Release 38 without specifying what version of the file was use. Gencode v38 is available with coordinated from GRCh37 and GRCh38. Please clarify what was used. If Gencode v38 was used with coordinated from GRCh38 they authors need to repeat their analyses matching mutations to genes (more on this on point 3).
2) In the same paragraph, the authors described a whole analysis that reports 75,756 genes. The human genome does not even have that many genes, so please explain where that number come from because there is no methods about it or I would not find it/connect with any other text.
3) According to the authors: "Among the 183,133 substitution mutations identified by Kucab et al. (2019), 92,204 occurred in known genes". Where does this information come from? I would not find a list of genes associated to the mutation in the original paper so I have to assume that was done by the authors. How? There is some text about identifying CDS but I can not tell if that refers to

the same analysis.
4) The authors "hypothesized that mutational patterns would be similar between these samples and the PAH-treated iPSCs", but they are comparing mutational rate from 1 iPSCs line treated in the lab to cancer cells from 14 different individuals exposed in natural conditions and with their different genetic background. The hypothesis makes no sense to me, since both systems are completely different. The results are, however, very much expected. Would please the authors explain what are the similarities in the two experiments that make the authors expect any other results?
5) The GO analyses sections are incredibly confusing, using wrong terminology and without methodology details that would aim a reader to understand what was done. First the authors claimed to have performed "gene ontology analysis on the 692 coding sequence (CDS) variants in PAH-treated iPSCs". "Sequence variants" is a term that mean very different things to different people, and it does not necessarily mean genes, which is what I hope the authors meant. If I am wrong and these "coding sequence (CDS) variants" are not genes, please explain what kind of enrichment this is. If I am right, please use the most common and not confusing term.
Assuming they actually did mean "genes" and performed a gene enrichment analysis, Where are the methods about how was this done? What method, or package and databases were used? What was the background genes used for the enrichment?
6) More of GO analysis: they performed an enrichment for 692 of the 769 disease genes. It is of no surprise that they found enrichment of functions related to the diseases they selected. This analysis makes no sense whatsoever, the starting point was disease genes. The analysis makes no sense unless they are comparing the enrichment of subsets of genes, e.g.: for one disease type, against all the other genes in the set or against all the genes in the genome. As it is right now the authors selected a set of disease genes, performed a GO enrichment and found that these are associated to biological functions associated to these diseases. There is no need to perform such an analysis to know what the results will be.
7) It is not possible to understand the value of the enrichment of T2D and obesity without reporting how many of the originally selected genes were used for the analysis because were mutated and in which conditions.
8) The "Genome-wide PAH adduct repair data come from translesion excision repair-sequencing (tXR-seq) of GM12878 cells". How are GM12878 cells comparable to iPSC?
9) It is not possible to understand figure S2 with the information provided in either the manuscript, the supplemental methods, or the text in the figure. What is the "mean mutations per subclone and mutations per gene per subclone"? In the main text, I suggest the authors also explain what they mean with "subclone" (page 5). Please add the text in the x-axis indicating what they represent.

Other comments:
- Where was the information about GC content come from? Gencode gives genomics location, not sequence. Please provide the details.
- Please make it clear in the text that mutation rates were derived from just one iPSC line. Right now the authors use the term "lines" implying they have more

one cell line.
- Supplemental method, page1: R is not a statistical package, but a software.
